



Machine Learning (ML)

Bem-vindo ao nosso estudo sobre Machine Learning. Este estudo ajudará você a compreender melhor como estudar a disciplina de Desenvolvimento de Algoritmos de Inteligência Artificial. Neste módulo será abordado sobre o que é Machine Learning, seus tipos, o que é o MNIST, o classificador do vizinho mais próximo e sobre regressão.
Conhecer os conceitos que serão apresentados aqui te ajudará a chegar ao sucesso do seu estudo. Vamos começar a estudar?

O que é Machine Learning?
O interesse pela Machine Learning aumentou exponencialmente na última década, as mudanças vistas no mundo onde cada vez mais é possível interagir com as aplicações dos computadores. Sistemas de recomendação, marcações de fotos de redes sociais, assistentes pessoais por voz, softwares para detectar spam, smartphones e smarttvs entre outros.
Mas o que seria Machine Learning - ML? Na tradução para português, Aprendizado de Máquina é a subárea da Inteligência Artificial que estuda métodos computacionais para adquirir novos conhecimentos, novas habilidades e novos meios de organizar o conhecimento já existente (VALDATI, 2020).
Pode-se dizer que os modelos utilizados no ML tem como objetivo fazer previsões a partir de outros conjuntos de dados através dos algoritmos existentes, cabendo ao Cientista de Dados escolher qual o melhor para cada situação ou problema a se resolver.
Não é fácil definir aprendizado, mas é muito importante para que seja possível desenvolver sistemas inteligentes capazes de aprender. Sendo assim, aprendizado é a capacidade de se adaptar, modificar ou melhorar o



comportamento e suas respostas, por isso uma das características mais importantes dos seres inteligentes (humanos ou não).

A ideia de Machine Learning é tentar reproduzir nas máquinas, o mesmo processo de aprendizagem dos humanos através dos algoritmos ML que utilizam Matemática e Estatística.

Tipos de Machine Learning

Normalmente os tipos de ML são divididos em subáreas de acordo com os tipos de problemas que deverão ser resolvidos.

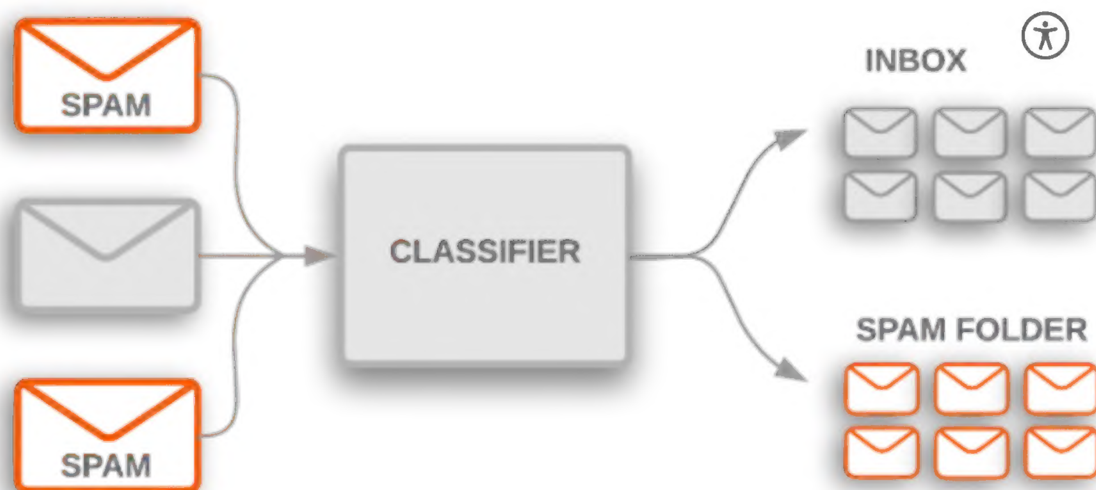
Aprendizado Supervisionado: através de uma entrada de dados com o valor correto (ou aproximadamente correto) o sistema deverá fornecer suas saídas desejadas. A cada exemplo o sistema deve aprender com as entradas e saídas e assim classificar novos entradas com base no que foi aprendido pelos exemplos.

Portanto, ao serem percebidas as entradas e saídas de qualquer situação é chamado de aprendizado de máquina supervisionado (NORVIG, 2003).

Algoritmos de classificação e regressão são exemplos de aprendizado supervisionado.

Classificação prevê classes pré-definidas e regressão prevê valores numéricos contínuos.

Alguns algoritmos de Classificação mais conhecidos: KNN, Naive Bayes, Logistic Regression, Support, Vector Machines, Decision Trees (TERRALAB,2022).



A Figura 1 mostra um exemplo de classificação de e-mails como spam. É possível perceber que os e-mails supostamente classificados como spam vão direto para uma caixa separada no aplicativo de e-mails. Como isso ocorre? Para que essa classificação ocorra um modelo de classificação baseado em entradas etiquetadas é usado. Essas entradas possuem e-mails classificados como suspeitos e outros confiáveis, sendo assim, o modelo aprende a conhecer essas classes e a identificar em qual classe o novo dado deverá ser rotulado (TERRALAB,2022).

Regressão é utilizada por exemplo para prever valores de imóveis ou a quantidade de produtos que serão vendidos em uma determinada época (TERRALAB,2022).





É possível observar na imagem acima que o Modelo de Regressão é dividido em simples e múltiplo, o que difere um do outro é o número de variáveis envolvidas. O Modelo de Regressão Simples envolve somente uma variável e o múltiplo duas ou mais. Além disso, para cada um dos tipos mostrado na imagem da Figura 2 ainda existe outra ramificação que divide esses modelos em lineares ou não lineares (RODRIGUES, 2018).
Alguns algoritmos de Regressão mais conhecidos: Linear Regression, Polynomial Regression, Logistic Regression, Principal Components Regression (PCR).

Aprendizado Não Supervisionado: é quando o algoritmo não tem um conjunto exemplo para o aprendizado com os resultados esperados. Essa abordagem aprende a prever através de suas percepções futuras considerando as percepções antigas (NORVIG, 2003). O conjunto de exemplos não é rotulado, portanto o sistema deve classificar esses conjuntos através das características semelhantes em seus dados de entrada.
Algoritmos de Agrupamento e Associação são exemplos de aprendizado não supervisionado.
No algoritmo de Agrupamento os dados de entrada são agrupados a partir de suas semelhanças ou diferenças. Exemplo: K-means.

E nos algoritmos que usam Regras de Associação os dados de entrada são agrupados de acordo com a associação de seus atributos. Exemplo: Apriori, Eclat e FP-Growth.

Aprendizado Semissupervisionado: é quando se utiliza técnicas de aprendizado supervisionado e não supervisionado. Nesse tipo de abordagem o modelo tem duas fases de treinamento, uma com dados com resultados conhecidos e outra com dados com resultados desconhecidos, conforme foi visto nos tópicos anteriores.

Alguns exemplos de uso: Diagnósticos médicos, detecção de fraudes online, sistemas de recomendação, reconhecimento da fala entre outros.

O que é MNIST?

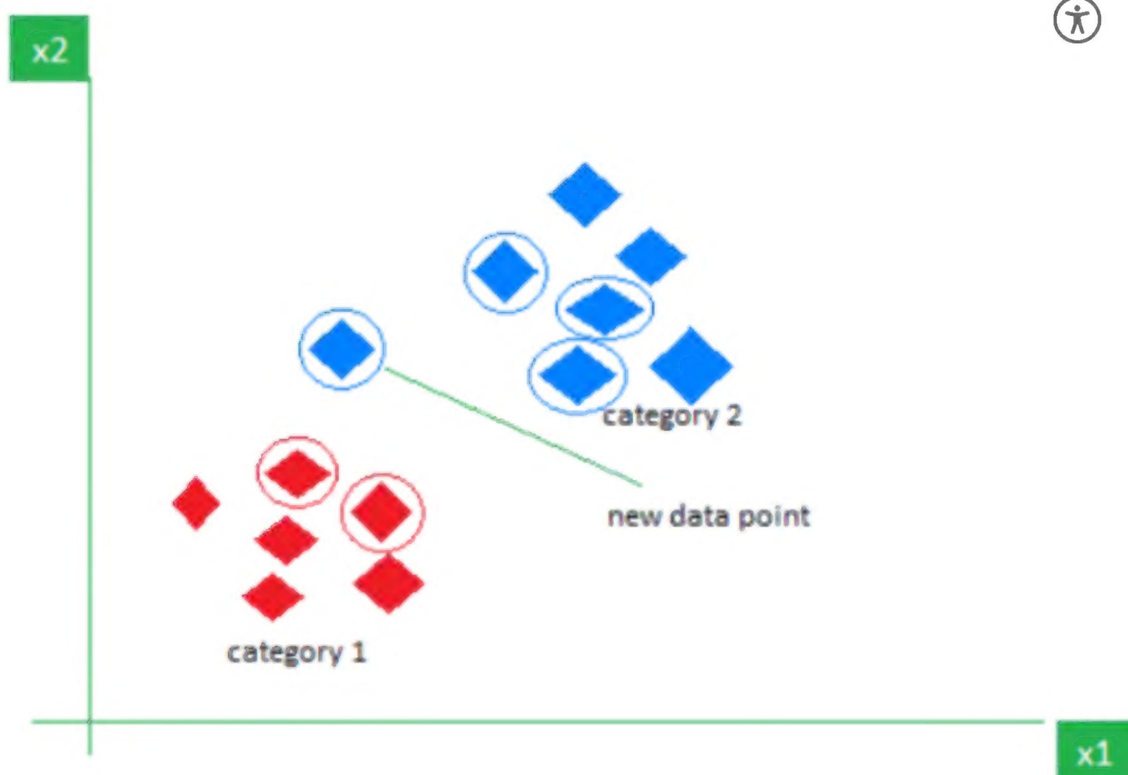
MINIST é um conjunto de dados de dígitos escritos à mão muito usado em estudos sobre Machine Learning. É gratuito e fica disponível no endereço eletrônico http://yann.lecun.com/exdb/mnist/ (acesso em 11/10/2022) para testes de treinamentos de algoritmos de Machine Learning.

Classificador do vizinho mais próximo

Um dos classificadores mais simples, também conhecido como K-nearest neighbors – K-nn. Esse classificador recebe um item para analisar e encontrar qual classe ele pertence, ou seja, com quais itens ele é mais parecido ou próximo.

A nomenclatura "vizinho mais próximo" refere-se à distância ou similaridade entre instâncias (características). Para calcular a proximidade é importante escolher um cálculo de distância (Distância de Manhattan ou Distância Euclidiana, por exemplo).





Na figura 3, é apresentado um conjunto de dados de treinamento, eles estão separados por classes (vermelha e azul). O losango azul circulado na imagem será usado como exemplo do funcionamento do método do vizinho mais próximo.

A cada iteração do algoritmo do K-nn, o losango é verificado, ou seja, se possui características parecidas com dados das duas classes existentes, após essa análise o losango é classificado como o vizinho mais próximo dos elementos da classe azul. Na figura 4 é apresentado os passos do algoritmo K-nn.

Teoria



No algoritmo KNN, K especifica o número de vizinhos e seu algoritmo é o seguinte:

- Escolha o número K do vizinho.
- Pegue o K vizinho mais próximo do ponto de dados desconhecido de acordo com a distância.
- Entre os vizinhos K, conte o número de pontos de dados em cada categoria.
- Atribua o novo ponto de dados a uma categoria, onde você contou a maioria dos vizinhos.

Atividade Extra

Exemplo: Para se aprofundar no assunto desta aula assista o filme "Free Guy - Assumindo o Controle" (2021)

Referência Bibliográfica

LUGER, George F. Inteligência Artificial. São Paulo: Pearson, 2013.

RICH, Elaine. Inteligência Artificial. São Paulo: McGraw-Hill, 1988.

RODRIGUES, Rodrigo Lins. Estatística Computacional. Modelos de Regressão Linear. Universidade Federal Rural de Pernambuco - UFRPE. 2018.

TERRALAB. Tipos de aprendizado de máquina e algumas aplicações. Disponível em http://www2.decom.ufop.br/terralab/tipos-de-aprendizado-de-maquina-e-algumas-aplicacoes/ Acesso em 04 de Outubro de 2022.

VALDATI, Aline de Brittos. Inteligência Artificial – IA. São Paulo: Contentus; 2020.



Ir para exercício